The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs

Katsutomo Okamura1, Wei-Jen Chung1, J. Graham Ruby2, Huili Guo2, David P. Bartel2 & Eric C. Lai1

In contrast to microRNAs and Piwi-associated RNAs, short interfering RNAs (siRNAs) are seemingly dispensable for host-directed gene regulation in Drosophila. This notion is based on the fact that mutants lacking the core siRNA-generating enzyme Dicer-2 or the predominant siRNA effector Argonaute 2 are viable, fertile and of relatively normal morphology1,2. Moreover, endogenous Drosophila siRNAs have not yet been identified. Here we report that siRNAs derived from long hairpin RNA genes (hpRNAs) programme Slicer complexes that can repress endogenous target transcripts. The Drosophila hpRNA pathway is a hybrid mechanism that combines canonical RNA interference factors (Dicer-2, Argonaute 2) with a canonical microRNA factor (Loquacious) to generate ~21-nucleotide siRNAs. These novel regulatory RNAs reveal unexpected complexity in the sorting of small RNAs, and open a window onto the biological usage of endogenous RNA interference in Drosophila.

Artificial, long-inverted repeat transcripts are efficiently processed by a Dicer-2 (Dcr-2)/Argonaute 2 (AGO2)-driven RNA interference (RNAi) pathway in transgenic Drosophila3,4. We hypothesized that this might reflect the existence of an endogenous pathway that accepts long, inverted repeat transcripts. To test this idea, we searched for inverted repeats using EINVERTED5 and selected putative hairpins containing mapped small RNA reads (see Methods). Out of 8,132 candidate regions, most consisted of the terminal inverted repeats of individual transposable elements or long terminal repeats of tandem inverted transposable elements. The remaining loci corresponded to inverted tandem duplications of messenger RNA- or transfer RNA-encoding genes, a microRNA (miRNA) gene (mir-997), a novel tandem pair of short hairpins (chou39-1 and chou39-2, Supplementary Fig. 1), and a handful of single-gene annotations and unannotated regions.

We analysed the size distribution of cloned RNAs from all of the non-transposable-element EINVERTED hits. Although these mostly exhibited a broad length distribution across the ~18–26-nucleotide cloning range, indicative of degradation fragments (Supplementary Fig. 1b), seven genomic regions specifically generated 21–22-nucleotide RNAs (Supplementary Fig. 1c–f). These included genes annotated as CG18854, CG32207, CR32205 and pncr009 (also known as pncr009:3L), a series of 20 repeats that partially overlap an intergenic region adjacent to CG4770 (Supplementary Figs 2–7). Except for CG4068, the coding potential of all of these loci is limited5,6. Still, we chose to introduce an ‘hp’ prefix to these six loci to distinguish the small-RNA-generating hairpins (‘hpRNAs’) from the potential protein-encoding segments of these transcripts.

The hpRNA hairpins were collectively much longer than typical animal pre-miRNAs, and several were even longer than plant miRNAs7. All hpRNA loci produced dominant small RNAs that presented duplexes with 2-nucleotide 3’ overhangs, implying RNase III processing (Fig. 1 and Supplementary Figs 2–7). Such an origin was more evident with the hp-CG18854 and hp-CG4068 hairpins, from which many consecutive, phased, small RNA duplexes were cloned (Fig. 1). The 20 tandem repeats at the hp-CG4068 locus were suggestive of local duplications, and created potential for a vast array of higher-order hairpin conformations (Fig. 1b and Supplementary Fig. 3). In addition, hp-CR32205, hp-pncr009 and hp-CG32207 were related in sequence and located within a 70-kb interval (Supplementary Figs 5–8). Thus, hpRNAs, like miRNAs, can apparently evolve as local genomic duplications.

We probed the consequences of dsRNA-mediated knockdown of candidate factors on hpRNA biogenesis. We first confirmed the potency of these knockdowns by analysing Bantam, the pre-miRNA and/or mature miRNA of which were sensitive to Drosha, Pasha, Dcr-1, Loquacious (Loqs), Exportin-5 (Exp5, also known as Ranbp21) and Argonaute 1 (AGO1), as expected (Fig. 2a). The behaviour of hp-CG4068B/D/G and hp-CG18854A contrasted sharply with that of bantam (Fig. 2a and Supplementary Fig. 10a). Consistent with their apparent derivation from phased cleavage of long inverted repeats, their processing was unaffected by Dcr-1 depletion, but was strongly dependent on Dcr-2. In addition, ~21-nucleotide hpRNA products were markedly reduced when AGO2 was depleted. Incidentally, the Dcr-2/AGO2-dependent accumulation of ~21-nucleotide (siRNA) and ~42-nucleotide (terminal loop) hp-CG4068D isoforms (Fig. 2a) provided evidence for the in vivo processing of both ‘single-repeat’ and ‘double-repeat’ (or higher-order) forms of the hp-CG4068 hairpins (Fig. 1b).

Several other aspects of hpRNA biogenesis deserve mention. First, we were surprised that hpRNA processing was very strongly dependent on the Dcr-1 cofactor Loqs. This was especially unexpected in light of the recent realization that the loqs null condition only mildly compromises the maturation of many miRNAs, such as Bantam (Fig. 2a). Second, mature hpRNA products declined reproductively in AGO1-deficient cells, which suggested the possible involvement of both AGO proteins in hpRNA biogenesis and/or function. Third, knockdown of Dcr-2, AGO2 and, to a lesser extent, AGO1 resulted in a ladder of hybridizing bands consistent with impaired hairpin processing (Fig. 2a and Supplementary Fig. 10a). This suggested that, in addition to Dcr-2, AGO proteins might also participate in hpRNA biogenesis. A role for AGO proteins has also been suggested for the maturation of siRNA duplexes and some pre-miRNAs9,10. Analysis of mutant animals corroborated this picture of hpRNA biogenesis, because mature hpRNA products were strongly reduced in Dcr-2, loqs and AGO2 homozygous mutants (Fig. 2b).

We next analysed the termini of hpRNA-derived small RNAs. β-elimination of RNAs with two free hydroxyl groups at their 3’ termini increases their mobility in denaturing polyacrylamide gel
electrophoresis, whereas treatment with calf intestinal phosphatase (CIP) reduces the mobility of 5’ monophosphorylated RNAs. Accordingly, miRNAs run faster after β-elimination and slower after CIP treatment (Fig. 2c and Supplementary Fig. 10b). CIP tests also indicated the presence of 5’ phosphates on hpRNA products (Supplementary Fig. 10b), but all of them were resistant to β-elimination indicating modification of the 3’ terminal ribose (Fig. 2c). Drosophila Hen1 methylates Piwi-associated RNAs (piRNAs) and exogenous siRNAs at their 3’ termini. We found that hen1 mutants exhibited lower levels of mature hpRNA products (Fig. 2b), and these were now fully sensitive to β-elimination (Fig. 2c). These data supported the classification of hpRNA products not as miRNAs, but as siRNAs.

We tested the regulatory activity of hpRNA-derived siRNAs using artificial luciferase transcripts linked to target sites that were complementary to various hp-CG4068- and hp-CG18854-derived siRNAs. Their activity was analysed in cells that overexpressed hp-CG4068 or hp-CG18854, with non-cognate pairs controlling for the generic effect of hpRNA overexpression. These tests revealed the specific repression of hp-CG4068B and hp-CG4068C sensors by single- and double-repeat hp-CG4068 expression constructs (Fig. 3a), and of the hp-CG18854B sensor by hp-CG18854 (Fig. 3b). However, a sensor for hp-CG4068D was not affected by ectopic hp-CG4068, consistent with its lower read count compared to hp-CG4068B and hp-CG4068C.

To address the activity of endogenous hpRNAs expressed by S2 cells, we asked whether 2’-O-methyl antisense oligonucleotides (ASOs) could derepress these sensors. Indeed, ASO-hp-CG18854B (but not ASOs to hp-CG18854A or hp-CG4068B) induced approximately twofold derepression of the hp-CG18854B sensor (Fig. 3c). Reciprocally, we observed that ASO-hp-CG4068B (but not other ASOs) resulted in a approximately twofold activation of the hp-CG4068B sensor (Fig. 3d). Thus, both exogenous and endogenous hpRNAs generate inhibitory RNAs. Some miRNAs are partially loaded into AGO2 (refs 16 and 17), but hpRNA products are the first endogenous Drosophila small RNAs known to be preferentially sorted to AGO2 as a class. This provided an opportunity to ask whether endogenously programmed AGO2 functions by means of slicing, translational repression, or both. We prepared hp-CG4068B sensors carrying tandem perfect sites, centrally bulged sites, or bulged plus seed mismatched sites. Both mutant sensors were strongly derepressed, and to roughly the same extent, relative to the perfect sensor (Fig. 3e). In fact, the activity of the

![Figure 1](https://example.com/figure1.png)  
**Figure 1** Examples of Drosophila hp-repeat transcripts. a, hp-CG18854 contains a >400-bp duplex separated by a large loop; the enlarged region highlights the phased nature of small RNA duplexes. Northern probes were designed against the RNAs labelled in blue. B, The hp-CG4068 locus consists of 20 tandem repeats that partially overlap the 3’ UTR of CG4068 (Supplementary Fig. 3) and generate phased small RNA duplexes. Each repeat adopts a hairpin structure, but higher-order hairpins are possible because repeats are complementary to each other; ‘1-repeat’ and ‘2-repeat’ isoforms are shown. Distinct small RNAs were cloned from related repeats with minor sequence differences (for example, RNAs highlighted in red and green).

![Figure 2](https://example.com/figure2.png)  
**Figure 2** Distinct biogenesis pathways for miRNAs and hpRNAs. a, Unlike miRNAs (for example, Bantam), hpRNA biogenesis in S2 cells is highly dependent on Dcr-2 and AGO2; Loqs and AGO1 suppression affect both miRNA and hpRNA biogenesis. b, miRNA and hpRNA biogenesis in pharate adult Drosophila. miR-8 was affected only in the loqs mutant, whereas hpRNA products were strongly decreased in the Dcr-2 and loqs mutants ( <10%), and significantly affected in the AGO2 and hen1 mutants. c, Modification of hpRNA-derived small RNAs is mediated by Hen1. Note that hpRNA products from hen1 mutants run as a range of faster-migrating species after β-elimination. CS, Canton S (a strain of fruitfly); KO, knockout deletion strain.
mutant sensors was comparable to the perfect sensor in the presence of cognate ASO (Fig. 3d). These data support the notion that hp-CG4068B exerts its major regulatory effect by slicing, with comparably little contribution from translational repression by AGO2.

With these functional regulatory data in hand, we searched for endogenous targets. We first considered whether hp-CG4068B might regulate the overlapping CG4068 3′ UTR (Supplementary Fig. 3), but gain- and loss-of-function tests were negative (Supplementary Fig. 11). Searches for trans-encoded targets were hindered by the fact that hp-CG4068 was only identifiable in the related species D. melanogaster, D. simulans and D. sechellia (Supplementary Fig. 3). However, hp-CG4068B, the most abundant siRNA product of hp-CG4068, contains 20 nucleotides of antisense complementarity (including three G-U pairs) to the coding region of mus308 (Fig. 3f). This transcript encodes a DNA polymerase/helicase required for DNA repair after exposure to crosslinking agents12. Compensatory co-variation between D. melanogaster and D. sechellia hp-CG4068B and mus308 target sites were suggestive of functional conservation (Fig. 3f). Consistent with this, we observed that mus308 levels were increased 2–3-fold in cells depleted of Dcr-2 or AGO2 (Supplementary Fig. 12), and that a luciferase–mus308 sensor was specifically repressed by single-repeat and double-repeat hp-CG4068 expression constructs (Fig. 3a).

Because our data indicated that hpRNAs generate functional siRNAs that are primarily dependent on AGO2, we tested whether endogenous hp-CG4068B complexes exhibited Slicer activity. Endogenously programmed complexes indeed cleaved a perfect endogenous target substrate in a manner that was competed away by ASO-hp-CG4068B (Fig. 3d). These data support the notion that hpRNAs can repress endogenous targets.

In plants, long hairpin RNAs from transgenes and long, extensively paired (and presumably very young) miRNA hairpins are substrates of DICER-LIKE4 (refs 19 and 20), and thus mature through a pathway distinct from that of canonical miRNA hairpins, which are substrates of DICER-LIKE1 (ref. 21). Likewise, we have shown that Drosophila hpRNAs enter a pathway distinct from that of miRNAs. Their derivation from unexpectedly long hairpins serves as an important caution for efforts to identify inverted-repeat small RNA genes. For example, some hpRNA-derived clones were recently reported but attributed incorrectly22, because only short genomic precursors were considered in that study.

We also searched for targets of hp-CG18854. The gene annotated as CG18854 is a possible pseudogene, because its open reading frame is short and poorly conserved1. CG18854 exhibits significant homology to the chromodomain gene CG8289 (Supplementary Fig. 13), and some of the abundant hp-CG18854-derived siRNAs exhibited extensive antisense complementarity to CG8289. When tested individually, siRNA-complementary sites from CG8289 did not mediate significant repression (data not shown). We therefore examined whether hp-CG18854 could regulate a translational fusion of CG8289 containing an extended complementary sequence. We transfected S2 cells with either tub-GFP or a tub-CG8289-GFP plasmid along with various hpRNA expression constructs, and observed that the CG8289 specifically repressed the accumulation of CG8289-GFP (Supplementary Fig. 13). These data demonstrate that hpRNA products can repress endogenous targets.
The *Drosophila* pathway combines canonical RNAi (Dcr-2, Hen1 andAGO2) and miRNA (Loq) biogenesis factors—a revelation that highlights the incomplete nature of our current understanding of small-RNA-sorting mechanisms. Together with concurrent studies that identify endogenous siRNAs from transposons and *cis*-natural antisense pairs in *Drosophila*22–24, this work sets the stage for directed studies of the genetic requirements for host-directed RNAi in this organism.

**METHODS SUMMARY**

We used EINVERTED2 to identify candidate genomic hairpins contained within 10-kb windows that satisfied a cutoff score ≥80 and had ≥70% pairing within the duplex region. Their expression as small RNAs was analysed using ten 454 libraries, a Solexa female head library23,24, and a new set of Solexa imaginal disc/brain library (NCBI-GEO GSM275691). We defined candidate hpRNA loci as non-transposon inverted repeats for which the duplex region generated more than three times as many 21–22-nucleotide RNAs than all other-sized RNAs combined (Supplementary Figs 2–9). For functional tests, we followed published protocols for soaking RNAi in S2 cells and northern blotting25 from knockdown samples or pharate adult flies. For sensor tests, four-copy-site targets (hp-CG4068B, hp-CG4068C and hp-CG4068D sensors) or a two-copy-site target (mus308 sensor) were prepared by inserting oligonucleotides into a modified KpnI site of tub-GFP sensor consisted of a CG8289 fragment from genomic DNA and cloned into the clone (CG18854 cDNA) and were cloned similarly. A CG8289:GFP translational clone (CG4068B, hp-CG4068C and hp-CG4068D sensors) or a two-copy-site target (mus308 sensor) were prepared by inserting oligonucleotides into a modified version of psiCHECK2 (ref. 25). For hpRNA expression constructs, one or two hp-CG4068 repeats were cloned into the 3′ UTR of UAS-DsRed; CG18854 fragments were amplified from genomic DNA (CG18854 genomic) or a D342473 clone (CG18854 cDNA) and were cloned similarly. A CG8289:GFP translational sensor consisted of a CG8289 fragment from genomic DNA and cloned into the KpnI site of tub-GFP26. RNA 5′ termini were analysed using periodate treatment in borax/boric-acid buffer followed by NaOH treatment (β-elimination) as described27. RNA 5′ termini were analysed using CIP as described28. For cleavage assays, we prepared labelled gel-purified templates using α-32P-GTP and capping enzyme (Ambion). Cleavage reactions were performed as described29 using S2-R+ cell extract. For detailed bioinformatic and molecular methods, see Methods.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 22 April 2008.

Published online 7 May 2008.


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We are grateful to R. Carthew, Q. Liu, H. Siomi and S. Cohen for plasmids and *Drosophila* strains. K.O. was supported by the Charles H. Revson Foundation. H.G. was supported by A*STAR, Singapore. D.P.B. is an HHMI investigator, and work in his laboratory was supported by a grant from the NIH (GM067031). E.C.L. was supported by grants from the Leukemia and Lymphoma Foundation, the Burroughs Wellcome Foundation, the V Foundation for Cancer Research, the Sidney Kimmel Foundation for Cancer Research, and the NIH (GM083300).

**Author Contributions** J.G.R. identified hp-CG4068 and hpRNA. W.-J.C. performed the EINVERTED analysis and identified the additional hpRNA loci and their targets. H.G. performed initial hpRNA northern analysis; all other experiments were designed and carried out by K.O. All authors contributed to the preparation of the manuscript.

**Author Information** The imaginal disc/brain sample described in paper has been deposited in the NCBI GEO under accession number GSM275691. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.C.L. (laie@mskcc.org).
We used EINVERTED using total density and cells miRNAs (Drosophila D773 probe (DNA), AATCAGCTTT- at 42 cells were resus- (refs 14 and 15). The sequences of the probes used in this microRNAs. buffer 3 (NEB, 100 mM NaCl, cells ml SDS–PAGE sample buffer. Western blotting was performed using rabbit proteinase K). RNA was recovered by phe- PLoS Biol. G Northern blotting. rifugation after 30 min incubation on ice. lumniter (Turner Biosystems). KpnI-CG8289 targetF, ggggtaccgccgccacc- nucotides for each well. Three days later, we lysed the cells and subjected them methyl antisense-mediated de-silencing assays (inhibitor sequences listed to generate dsRNA, and soaking RNAi was performed as described (NEB); other plasmids were described previously the indicated primers and were cloned into XhoI-XbaI sites of Litmus28i vector. hpRNA candidates (Supplementary Figs 2–9). UAS-DsRed-hpRNA plasmids into 1 forming quadruplicate transfections of 25 ng target, 12.5 ng ub-Gal4 and 25 ng UAS-DsRed-hpRNA plasmids were transfected to 2 × 105 cells in 6-well format. Three days later, transfected cells were harvested and lysed with 2× SDS–PAGE sample buffer. Western blotting was performed using rabbit anti-GFP (Molecular Probes) or mouse anti-α-tubulin (DM1A, Sigma).

For the green fluorescent protein (GFP) sensor assay, 250 ng target, 125 ng ub-Gal4 and 250 ng UAS-DSRed-hpRNA plasmids were transfected to 2 × 105 cells in 6-well format. Three days later, transfected cells were harvested and lysed with 2× SDS–PAGE sample buffer. Western blotting was performed using rabbit anti-GFP (Molecular Probes) or mouse anti-α-tubulin (DM1A, Sigma).

In vitro cleavage assay. Templates for in vitro cleavage targets were prepared by treating annealed oligonucleotides with Taq polymerase. The oligonucleotides A and B or A and C (below) were used for the template preparation for hp-CG4068-B target or mus088 target, respectively. Target RNAs were in vitro transcribed using Megascid T7 kit (Amplification) and purified by acrylamide gel electrophoresis. Purified RNA was labelled by 25-P-GTP using capping enzyme (Amplification) according to the manufacturer’s instructions. The cleavage reaction was performed as described5 using S2-R+ cell extract. S2-R+ cells were resus- pended in hypotonic buffer (30 mM HEPES-KOH, pH 7.4, 2 mM magnesium acetate, 5 mM DTT, 1 Complete mini EDTA free (Roche)) and lysed by five passages through a 25-gauge needle. The lysate was cleared by a centrifugation for 25 min at 14,600 g at 4°C, and was flash-frozen in 10-μl aliquots. Approximately 2,000 counts per min cap labelled RNA was incubated in a reaction mixture (50% S2 lysate, 0.5 mM ATP, 5 mM DTT, 100 mM KοAc, 0.1 μM 51 RNaseOut (Invitrogen)) for 1 h at room temperature. 2′-O-methyl- ASO inhibitors (Integrated DNA Technologies) were added at 100 nM concentra- tion and preincubated with the reaction mixture at room temperature for 20 min before the addition of the cap-labelled target RNA. Reactions were stopped by the addition of stop buffer (50 mM sodium chloride, 50 mM EDTA, 1% SDS, and 100 μM 1 proteinase K). RNA was recovered by phe- nol/chloroform extraction and ethanol precipitation. A, LucLett_3_region AS, attacactcatGAGGAGAAGGGACTTGGTTAAGTTGCGAGGCGGTG; B, T7 Hwy-CG40682B_Luc_sense, TAATACGACTCACTATAGGGAGCAGAGCTTTGAAGGGTGAATTC; C, T7 Mus088_Luc sense, TAAATACGACTCACTATAGGGGAGCCGTTTGAGTTGAGTCAGACTGACAT; D, T7 Mus088_Luc sense, TAAATACGACTCACTATAGGGGAGCCGTTTGAGTTGAGTCAGACTGACAT.

For the green fluorescent protein (GFP) sensor assay, 250 ng target, 125 ng ub-Gal4 and 250 ng UAS-DSRed-hpRNA plasmids were transfected to 2 × 105 cells in 6-well format. Three days later, transfected cells were harvested and lysed with 2× SDS–PAGE sample buffer. Western blotting was performed using rabbit anti-GFP (Molecular Probes) or mouse anti-α-tubulin (DM1A, Sigma).

In vitro cleavage assay. Templates for in vitro cleavage targets were prepared by treating annealed oligonucleotides with Taq polymerase. The oligonucleotides A and B or A and C (below) were used for the template preparation for hp-CG4068-B target or mus088 target, respectively. Target RNAs were in vitro transcribed using Megascid T7 kit (Amplification) and purified by acrylamide gel electrophoresis. Purified RNA was labelled by 25-P-GTP using capping enzyme (Amplification) according to the manufacturer’s instructions. The cleavage reaction was performed as described5 using S2-R+ cell extract. S2-R+ cells were resus- pended in hypotonic buffer (30 mM HEPES-KOH, pH 7.4, 2 mM magnesium acetate, 5 mM DTT, 1 Complete mini EDTA free (Roche)) and lysed by five passages through a 25-gauge needle. The lysate was cleared by a centrifugation for 25 min at 14,600 g at 4°C, and was flash-frozen in 10-μl aliquots. Approximately 2,000 counts per min cap labelled RNA was incubated in a reaction mixture (50% S2 lysate, 0.5 mM ATP, 5 mM DTT, 100 mM KοAc, 0.1 μM 51 RNaseOut (Invitrogen)) for 1 h at room temperature. 2′-O-methyl- ASO inhibitors (Integrated DNA Technologies) were added at 100 nM concentra- tion and preincubated with the reaction mixture at room temperature for 20 min before the addition of the cap-labelled target RNA. Reactions were stopped by the addition of stop buffer (50 mM sodium chloride, 50 mM EDTA, 1% SDS, and 100 μM 1 proteinase K). RNA was recovered by phe- nol/chloroform extraction and ethanol precipitation. A, LucLett_3_region AS, attacactcatGAGGAGAAGGGACTTGGTTAAGTTGCGAGGCGGTG; B, T7 Hwy-CG40682B_Luc_sense, TAATACGACTCACTATAGGGAGCAGAGCTTTGAAGGGTGAATTC; C, T7 Mus088_Luc sense, TAAATACGACTCACTATAGGGGAGCCGTTTGAGTTGAGTCAGACTGACAT.